

# 磁気テープストレージの 大量データアーカイブ高速化技術

吉田 武俊      飯澤 健      田村 雅寿

---

## あらまし

従来、磁気テープストレージはバックアップ用途としての利用が中心であった。その大容量性と低コスト性に加えて、近年の転送速度の高速化や、磁気テープ内のデータをファイル単位で取り扱うことができるテープ向けファイルシステムLTFS（Linear Tape File System）の普及によって、アーカイブ用途での活用が期待されている。しかし、磁気テープストレージは、テープ上の連続した領域を順にアクセスする性能は高いが、不連続な位置にランダムにアクセスする性能は低く、大量データのアーカイブ用途として利用するためには、ランダムアクセスの性能向上が求められていた。富士通研究所は、LTFSの機能拡張と、複数のテープカートリッジの仮想統合、テープ特性に合わせたデータ管理、およびアクセス順序制御をすることで、テープへのランダムアクセス性能を向上させた。

本稿では、開発した磁気テープストレージの大量データアーカイブ高速化技術について述べる。

---

## 1. まえがき

近年、ビッグデータやAIの活用によって、扱われるデータの形式（動画、画像など）やサイズが多様化するとともに、扱うデータの数も増えている。更に、その全体量もペタバイト級になってきている。このようなデータの管理には、保存データのサイズや数に制限がなく利便性が高いオブジェクトストレージが広く利用されている。オブジェクトストレージは多くのクラウド事業者でも提供されており、Amazonが提供するAmazon Web Services (AWS) のS3や、Microsoftが提供するAzure Blob Storage などがある。更に、証跡としての活用や、過去から今後のトレンド分析などのために、大容量のデータを数年から10年を超える長期間に渡って保存し続けるニーズが高まっている [1]。そのため、AWSのS3 Glacier [2] やDeep Archive, Microsoft AzureのAzure Archive Storage [3] などのストレージサービスのよう、アクセス性能を抑え容量あたりの保管コストを低くした大容量アーカイブストレージに注目が集まっている [4]。

しかし、クラウドにおけるアーカイブストレージサービスでは、保管コストは低いものの、データの取り出しに高額な料金が発生する場合がある。今後、企業のデジタルトランスフォーメーション (DX) の進展によって、様々な現場で発生する大量データの利活用が本格化し、更なるデータ増加が見込まれている。そういった大量データのアーカイブストレージとして、低コストで大容量な磁気テープストレージが再注目されている。

このような状況において富士通研究所では、大量データアーカイブに向けた磁気テープストレージのアクセス高速化技術を開発した。本稿では、開発

した仮想統合ファイルシステムとその高速化技術、および性能評価について述べる。

## 2. 磁気テープストレージの概要

大量データの利活用においては、時間変化や異なる分野のデータの関連性を分析するために、大量データを安価かつ長期的に保存できるストレージが望まれている。このような用途において、磁気テープストレージは他のストレージメディアと比較して優れた点がある。例えば、同サイズのデータを保持するために必要な電力量は、ハードディスクドライブ (HDD) のそれに比べて数十分の一 [5] である。また、LTO (Linear Tape Open) [6] として規格が統一されており、テープ容量は2年から3年ごとにほぼ2倍に増加している。表-1に、LTO規格におけるテープ容量および転送性能を示す。

テープストレージをアプリケーションから利用するためには、LTFS (Linear Tape File System) と呼ばれるソフトウェアが用いられる。LTFSは、テープ内のデータをHDDと同じようにファイル単位で取り扱えるようにするもので、LTO規格のLTO5からサポートされた。LTFSのフォーマットやAPI (Application Programming Interface) の仕様は2010年に公開されてデファクトスタンダードとなり、2016年4月に国際規格 (ISO/IEC 20919:2016) として標準化されている [5]。

LTFSのフォーマットは、テープのパーティショニングを用いて、インデックス・パーティションとデータ・パーティションの二つのパーティションから構成される。インデックス・パーティションには、データのファイル名、ファイルの操作（書き込み、削除、更新）日時、テープ上の記録位置などといったインデックス情報が記録され、テープの先頭部分

表-1 LTO規格の遷移

	2010年	2012年	2015年	2017年	2019年以降		
規格	LTO5	LTO6	LTO7	LTO8	LTO9	...	LTO12
容量 (非圧縮)	1.5 TB	2.5 TB	6 TB	12 TB	24 TB	...	192 TB
転送性能	140 MB/s	160 MB/s	300 MB/s	300 MB/s (HH) 360 MB/s (FH)	—	...	—

HH (Half-Height), FH (Full-Height) はテープドライブの厚さの規格

の小さな領域で管理されている [5]。データ・パーティションは、データの実体が記録される大きな領域であり、データを追記方式で書き込むとともに、定期的にインデックス情報の書き込みも行われる。

データの削除は、インデックス情報として削除するだけで、データ・パーティション上のデータは削除しない。そのため、削除したデータが存在していたテープ上の領域は、カートリッジを再フォーマットしなければ再び利用することはできない。また、データの更新は、新たなデータをデータ・パーティションに追記し、インデックス情報を更新する。この特性から、テープストレージは頻繁に更新されるデータの保存よりも、更新が少ないデータの保存が中心となる。

### 3. 従来技術の問題点

本章では、従来のLTFSに関する二つの問題について述べる。

#### (1) ランダム読み出し性能の低下

磁気テープはデータに連続的にアクセスする性能は高く、最新の磁気テープ規格であるLTO8においては360 MB/sに達する。その一方で、特に不連続な位置のデータをランダムに読み出す性能は低く、例えば数MB単位のデータに対するランダムな読み出しでは、連続読み出しの10分の1以下の性能になってしまう。

また、アクセスするデータのサイズが小さくなるほど、データを読み出す際の磁気ヘッドの位置合わせの処理が多くなり、アクセス性能が低下する。例えば、磁気テープ上にアクセス対象となる複数の小さいデータが、間に不要なデータを挟みながら断続的に記録されている場合、磁気テープの走行方向に断続的に読み出そうとすると、アクセス性能は大きく低下する。

磁気テープストレージをバックアップ用途で利用する場合は、まとまった単位での連続アクセスとなるため問題はない。しかし、必要なタイミングに必要なデータ単位でアクセスするアーカイブ用途で利用する場合は、様々なデータサイズにおける読み書き性能や、ランダムな読み出し性能の向上が必要となっていた。

#### (2) ファイル数増加による書き込み性能の低下

2章で述べたように、LTFSではファイルごとのインデックス情報を保持している。アーカイブ用途で磁気テープを使う場合、利用者は様々なサイズのファイルを使用することになるが、小さいサイズのファイルを大量に書き込むと、このインデックス情報も増大し、書き込み性能低下の原因となっていた。

## 4. 開発技術

前章で述べた問題に対して、富士通研究所ではLTFSの上位に複数のテープカートリッジを仮想統合するファイルシステムと磁気テープアクセス高速化技術を開発した (図-1)。この仮想統合ファイルシステムが複数のテープカートリッジを一つに仮想的に統合することで、利用者はそれぞれのテープカートリッジを意識することなく、必要なファイルにアクセスすることが可能となる。本章では、開発した磁気テープアクセス高速化技術について述べる。

#### (1) 物理位置を考慮したアクセス順序制御技術

磁気テープでは、テープの幅をラップと呼ばれる細い領域に分割し、ラップの端まで書き込んだら折り返して次のラップに書き込んでいく。このラップ上を折り返しながらデータにアクセスする場合のデータ位置は、テープ上の物理的な位置 (物理アドレス) とは別に抽象化された論理アドレスで管理される。二つのデータについて論理アドレスの差と物理アドレスの差を比べた場合、論理アドレスでは大きく離れている一方で、物理アドレスでは近い位置にある場合が考えられる (図-2)。

そこで、仮想統合ファイルシステムにおいて、複数の読み出し要求を受け入れた上で、論理アドレスではなく、テープ上の物理アドレスが近いものから順に処理していく方式を開発した。図-2の例において、論理アドレスによる管理では、同じラップ上にあるデータをテープの走行方向に準じて読み出す。これに対して、物理アドレスによる管理では、現在のヘッド位置に近い異なるラップ上のデータを読み出すこととなり、ヘッドの移動距離は最小限となる。

また、磁気テープへのアクセス時は、開始位置にヘッドを合わせる処理に時間を要する。そこで、

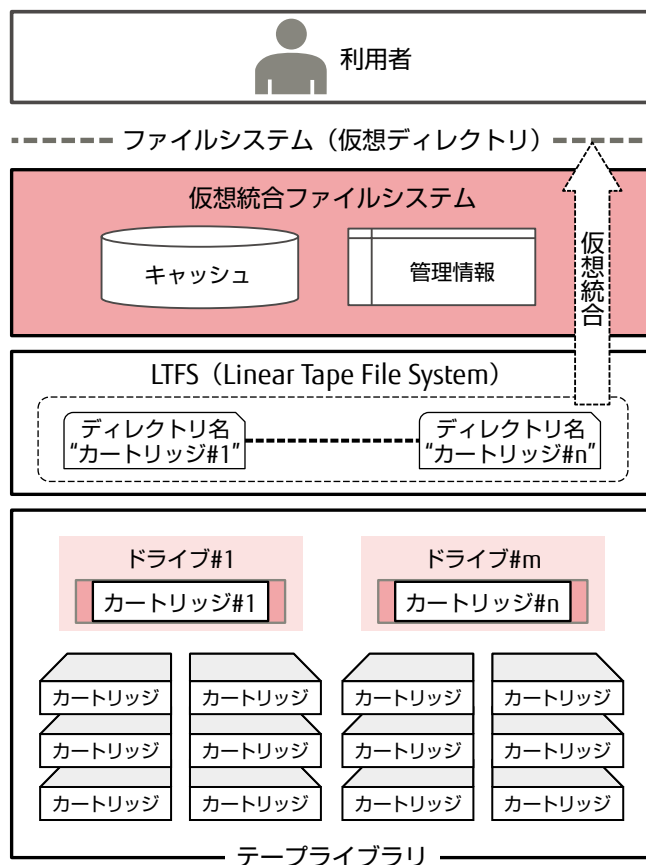


図-1 開発システムの構成イメージ

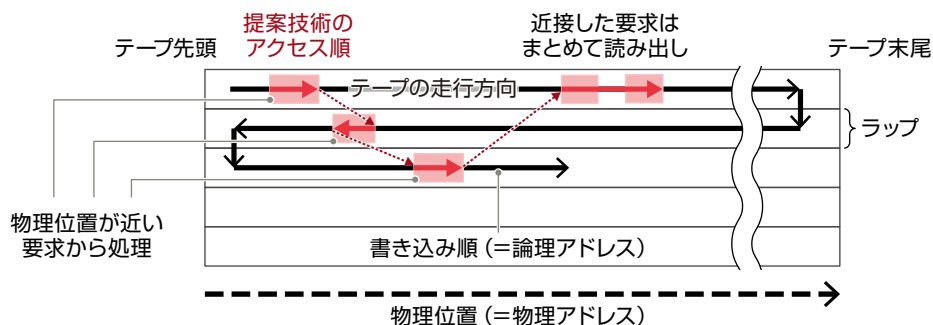


図-2 物理位置を考慮したアクセス順序制御

同一ラップ上で近接している二つのファイルに読み出し要求があった場合には、それぞれのファイルを断続的に読み出すのではなく、間に含まれるファイル群もまとめて読み出した上で、不要なファイルを廃棄することで更なる高速化を実現している。

## (2) 複数ファイルの集約管理技術

利用者からは様々なサイズのファイルが書き込まれるが、指定したサイズ以下の小さいファイルにつ

いては、利用者には小さいファイルとして見せつつ、LTFS上ではまとめて大きなファイルとして保持する仕組みを開発した（図-3）。これによりLTFSが管理するファイル数を軽減し、性能の低下を抑制している。

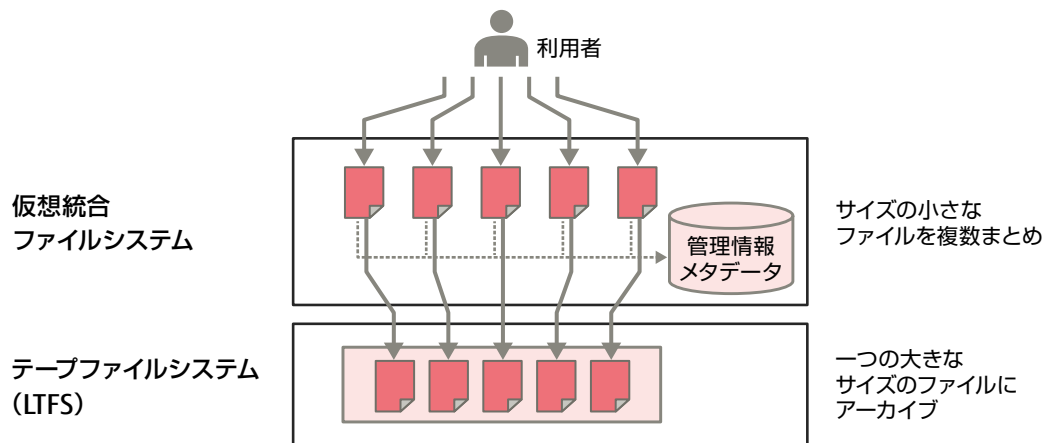


図-3 複数ファイル集約機能

## 5. 性能評価

本章では、本技術の性能評価について述べる。

オープンソースの分散ストレージソフトウェアであるCeph（セフ）[7]を利用して、HDDで構成されるストレージ層と磁気テープストレージで構成されるストレージ層の二つの階層ストレージを構築し、本技術を用いた場合のアクセス性能の評価を行った。

### (1) 物理位置によるアクセス順序制御の評価

磁気テープ上に蓄積した100 MBのサイズのファイル50,000個から、ランダムに100個のファイルを読み出した場合の所要時間の評価を行った。従来方式では5,400秒を要していたところ、本技術を用いることで、従来の約1/4となる1,300秒での読み出しを確認した。

### (2) 複数ファイル集約管理の評価

HDD上にある1 MBのサイズのファイル256個を磁気テープ上に階層移動させるのに要する時間の評価を行った。従来方式では2.5秒を要していたところ、本技術を用いることで、従来比約1/2となる1.3秒でのファイル移動を確認した。

本技術によって、アーカイブ用途で発生するランダム読み出しや多様なサイズをもつファイルの書き込みといった磁気テープアクセス性能の高速化が可能となり、大量データを長期に渡ってアーカイブする際に、価格対性能比に優れたデータアーカイブ基盤の実現が期待できる。

## 6. むすび

本稿では、磁気テープストレージの大量データアーカイブ高速化技術について述べた。仮想統合ファイルシステム技術、物理位置を考慮したアクセス順序制御技術、複数ファイル集約技術によって、爆発的に増加する大量データのアーカイブ用途として磁気テープストレージの活用を促進し、価格対性能比に優れたデータアーカイブ基盤を実現していく。

今後は、本技術の業務適用を想定した検証を進め、多様なアクセスにおける性能向上に対応していく予定である。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

### 参考文献・注記

- [1] Bill Kleyman : “Understanding Cold Storage: Best Practices around Google Nearline & Amazon Glacier”, CloudBerry Lab, Aug, 2015.
- [2] Amazon S3 Glacier.  
<https://aws.amazon.com/jp/glacier/>
- [3] Microsoft : Azure Archive Storage.  
<https://azure.microsoft.com/ja-jp/services/storage/archive/>
- [4] 藤巻秀明 他：コールドストレージを適用したビッグデータ解析基盤. FUJITSU, Vol. 66, No. 4, p. 49-54 (2015).

<http://img.jp.fujitsu.com/downloads/jp/jmag/vol66-4/paper07.pdf>

- [5] 一般社団法人 電子情報技術産業協会 テープストレージ専門委員会：テープストレージ動向<2018年版>.  
July, 2018.

[https://home.jeita.or.jp/upload\\_file/20180725112701\\_jYahZ5OUKI.pdf](https://home.jeita.or.jp/upload_file/20180725112701_jYahZ5OUKI.pdf)

- [6] Ultrium LTO.

<https://www.lto.org/>

- [7] Ceph.

<https://ceph.io/>

この記事は、富士通の技術情報メディア「富士通テクニカルレビュー」に掲載されたものです。他の記事も是非ご覧ください。

富士通テクニカルレビュー

<https://www.fujitsu.com/jp/technicalreview/>



## 著者紹介



### 吉田 武俊 (よしだ たけとし)

株式会社富士通研究所  
ICTシステム研究所  
データマネジメントシステムの研究に従事。



### 飯澤 健 (いゐざわ けん)

株式会社富士通研究所  
プラットフォーム革新プロジェクト  
データマネジメントシステムの研究に従事。



### 田村 雅寿 (たむら まさひさ)

株式会社富士通研究所  
プラットフォーム革新プロジェクト  
データマネジメントシステムの研究に従事。